Project Methodology

Team Members:

- Andrew Dunn, andrewdunn@uchicago.edu
- Gregory Ho, gregoryh@uchicago.edu
- Santiago Segovia Baquero, ssegovba@uchicago.edu
- Stephania Tello Zamudio, stello@uchicago.edu

Introduction

Evictions are very impactful events, but their relationship with other factors in people's lives would benefit from further research. This project looks to understand further the neighborhood characteristics associated with eviction in the city of Chicago. Using policing data, rental costs, median income, and distance calculations, we construct an index to measure neighborhood deprivation, using a similar approach as the Multi-dimensional Poverty Index. The resulting maps and analyses show that evictions are highly correlated with the deprivation index we create.

Specifically, we make a correlational map of evictions and the constructed deprivation index, as well as graphs 1) correlating the index and the individual indicators which compose it with eviction and 2) comparing zip codes in the city of Chicago to the city average across these indicators.

Data Sources

The data sources used for this project and the relevant variables we used are described below. Our geographic level of analysis was zip codes in Chicago.

- ACS Data: We obtained total population figures and median income from the 2019 ACS.
- Zillow Data: We obtained data from the Zillow Observed Rent Index (ZORI), which is a smoothed measure of the typical observed market rate rent across a given region. The data is available on a monthly basis from 2015 and the level of disaggregation is at a zip code level. We took the average of all months in 2019 to obtain average rent figures.
- Evictions Data: Evictions data was sourced from the *Law Center for Better Housing* (LCBH)[1] . The data was filtered by year (2019).
- Crime Data: We accessed crime data in the year 2019 from the Chicago Data Portal API.
- Google Maps API: We accessed this API to calculate the time and distance between a series of addresses.

Data Cleaning

This process involved working on all the different data sources described in the previous section. After the data were accessed via API, they were loaded as pandas DataFrames, and manipulated in order to create a merged analysis file. One zip code, 60666, which contains solely O'Hare airport, was excluded due to a lack of data. The columns in the merged analysis file correspond to the characteristics described as follows:

---

[1] https://eviction.lcbh.org/

- Geospatial Boundaries:

  We obtained two geospatial files containing boundaries for zip codes and census tract from the Chicago Data Portal.[2]

- ACS Data:

  The data corresponds to the 2019 Census. The only filter applied to this dataset was to select the corresponding zip codes for the city of Chicago.

- Zillow Data:

  The data was filtered by year (2019) and state (Illinois). The only zip codes considered are the ones corresponding to Chicago. Given that the rent data has a monthly frequency, a mean rent price is calculated at a zip code level. Some zip codes in Chicago don't have rent data in Zillow. For these cases, a median rent is calculated based on the other zip codes and the corresponding value is imputed for the missing observations.

- Evictions Data:

  Evictions data was sourced from the *Law Center for Better Housing* (LCBH). The data was filtered by year (2019), and a mapping between the census tract and zip codes was performed using geopandas. The data is aggregated at a zip code level.

- Crime Data:

  The first step was to define crime categories, our classification of crime was adapted based on the FBI's Uniform Crime Reporting Handbook. The FBI employs two broad classifications - Part I offenses (criminal homicide, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, arson), and Part II offenses representing all other reportable classifications outside Part I (these range from fraud, vandalism, drug abuse violations, driving under influence, disorderly conduct, curfew and loitering laws, etc.).

  We adapted the classification of crime to be categorized as 'non-offensive or nonviolent crime', and 'violent crime'. The first category is defined as crimes that do not involve a threat of harm or an actual attack upon a victim. The second category involves offenses that directly affect a victim by force or threat of force.

  The crime data contained latitude and longitude coordinates, but not zip codes. We built a function using geopandas to merge zip codes with each row of this data using the latitude and longitudes. Given the size of the data, this function takes nearly 40 hours to run. To speed up the process, we created a .txt file with the output of that function, containing a dictionary-like format that maps each latitude-longitude pair to its respective zip code. This file is loaded in the main cleaning function to avoid repeating the process. Finally, the data is aggregated at a zip code level.

- Travel Data:

  For each zip code, we randomized a set of random coordinates bounded within each zip code boundary. For each coordinate, we obtained time taken and distance travelled by "private vehicle" to the Central Business District ("The Loop, Chicago"). These datapoints were aggregated at the zip code level. Travel data was extracted on Feb-2023. As Google

[2] https://data.cityofchicago.org/

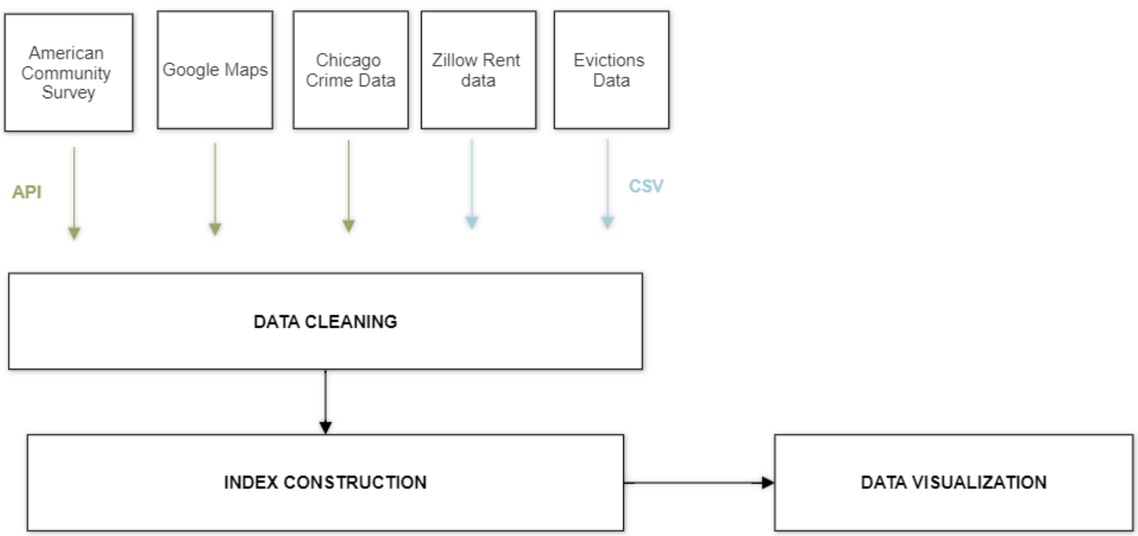
Distance Matrix API does not permit API calls on past travel data, we assume that distance and time is similar to 2019.

The final variables in the merged database (clean_database.csv), as well as their descriptions, are described in the following table:

Name of the variable	Description
crime_scaled	Number of crimes scaled by the population of their respective zip code.
violent_crime_scaled	Number of violent crimes scaled by the population of their respective zip code.
non_offensive_crime_scaled	Number of non-offensive crimes scaled by the population of their respective zip code.
RentPrice	Average of observed market rent in a specific zip code.
hh_median_income	Median income of households in a specific zip code.
eviction_filings_completed	Total number of evictions registered in a specific zip code.

Data Processing Pipeline

The graphic below shows, at a high level, the process we followed from data collection to data visualization.



Index Construction

To create a deprivation index, the project employed Sen's capability approach, a moral framework that assesses social arrangements based on individuals' freedoms to pursue the lives they have reason to value.[3] In Sen's framework, commodities refer to resources that individuals have access to (money, education, healthcare, etc.). These *commodities* enable individuals to pursue their desired *functionings* which are the various activities and achievements that people might have reason to value – these could be abstract ideas like meaningful work or having happiness.

One's place of residence plays an important role in shaping well-being. This is because one's access to resources and opportunities are bounded to the type and level of quality of healthcare, education, employment, and many other factors that constrains well-being. As a result, people living in deprived neighborhoods may face significant obstacles in pursuing their desired *functionings* and achieving a good quality of life.

In this project, we attempt to quantify factors that constraints level of well-being. We began by conducting a literature review[4] on what some of these components may be and how they are measured. In some cases (*rental affordability, time to central business district*), we were able to define deprivation thresholds based on the literature review, while in others, authors make no mention of thresholds (for example, how much crime is necessary for a neighborhood to be considered unsafe). In those cases, thresholds were determined relative to the 1st quintile (this is a rule of thumb measure).

Table 1 below summarizes the dimensions, deprivation thresholds, and studies which advocate each's use.

Table 1: Dimensions of Deprivation Index

Index	Sub-index	Threshold	Data Source
Safety[5]	Violent Crime	Incidence > 1st quintile	Chicago PD
	Non-offensive crime	Incidence > 1st quintile	Chicago PD
Housing affordability[6]	Rental affordability	Median Rent > 30% of median monthly income	ACS, Zillow
Transport Accessibility[7]	Distance to CBD	Travel distance > 1st quintile	Google Maps API
	Time to CBD	Takes more than 20m to travel to CBD	Google Maps API

[3] Sen, Amartya (1999). Development as freedom (1st ed.). New York: Oxford University Press. ISBN 9780198297581.

[4] Mouratidis (2020), Clark (2020)

[5] Our classification of crime was adapted based on the FBI's Uniform Crime Reporting Handbook. The FBI employs two broad classifications - Part I offenses (criminal homicide, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor vehicle theft, arson), and Part II offenses representing all other reportable classifications outside Part I (these range from fraud, vandalism, drug abuse violations, driving under influence, disorderly conduct, curfew and loitering laws, etc.)

[6] UN Habitat constructed a Global Sample of Cities (made out of 200 cities) statistically representing World Cities. Median rent affordability is 30% of household income. Standard rent affordability is 25% of household income.

[7] Travel time thresholds were determined based on the 20m neighborhood criterion - The ability to meet most of your everyday needs locally within a 20-minute journey from home. Source: World Economic Forum. "Making affordable housing a reality in cities." Cities, Urban Development and Urban Services Platform in Collaboration with Pw (2019).

Brief Description of the Alkire Foster Method

The starting point of constructing a Multidimensional Poverty Indicator would be the generation of a matrix representing zip codes and indicators of deprivation. To construct it, we pulled crime data from Chicago Data Portal, rental data from the Zillow API, income data from the US Census API, and travel data from Google's Distance Matrix API.:

$$Y = \begin{bmatrix} 100 & 3.4 & 280 & 800 \\ 250 & 4 & 450 & 1000 \\ 480 & 4.3 & 850 & 2800 \\ 640 & 5 & 1400 & 3200 \end{bmatrix}$$

This is a toy example of Matrix Y. Each row in matrix Y represents a zip code, while a column represents a variable of interest (sub-index in Table 1).

The AF method requires the specification of a deprivation vector, representing the standards[8] that are employed to label each analytical unit (zip code) as being deprived, or not in a particular dimension. The deprivation vector is expressed as vector z:

$$Z = [\quad 300 \quad 3 \quad 800 \quad 1200 \quad]$$

Each element in the deprivation vector represents thresholds, above which zip codes are labeled deprived or not deprived depending on whether the zip code's value falls above or below the threshold. Based on these set of standards, the next step is to construct Matrix G0, which is a binary matrix representing the following relationship:

$$G0(e_{i,j}) = \begin{cases} 1 \; if \; Y(e_{i,j}) < Z(e_{i,j}) \\ \quad 0 \; otherwise \end{cases}$$

Applying these thresholds from z on matrix Y yields the Deprivation Matrix (G0):

$$G0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

In the AF method, Alkire and Foster (2011) introduces a fixed cutoff k that is used to censor non-poor9 analytical units out of their analysis. In this step, for any analytical unit found with less than k deprivations, all elements of that row are assigned 0.

The next step in the method is to compute adjusted poverty gap (Matrix G1). Matrix G1 re-expresses each column in G0, based on Matrix Y in terms of their distances from the deprivation vector z. In this way, matrix G0 is often preferred because it satisfies monotonicity and is sensitive

---

[8] Sen's framework requires that components of the index and their weights should be decided by a democratic process or be made public that public debates over its dimensions and weights might take place. In policy, different localities tend to adapt this deprivation vector (representing the set of standards they employ) according to their localities qualitative preferences and beliefs. Sen's capability approach is incomplete by design, allowing the basis for index construction to not be predetermined a priori.

[9] Alkire and Foster (2011) explicitly distinguishes households that are poor vs. households that face deprivation in particular dimensions. We do not need to make such a distinction at this point in time, hence we adopt a k=0 (no cutoff) in our analysis.

to both the breadth and depth of individual deprivations. Each element in matrix G1 is defined as follows:

$$f(e_{i,j}) = \frac{Y(e_{i,j}) - Z(e_{i,j})}{Z(e_{i,j})}$$

An example of G1 is as follows:

$$\text{G1} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.33 & 0.00 & 0.00 \\ 0.60 & 0.43 & 0.06 & 1.33 \\ 1.13 & 0.67 & 0.75 & 1.67 \end{bmatrix} \begin{matrix} 0.00 \\ 0.33 \\ 2.42 \\ 4.22 \end{matrix}$$

Since, matrix G1 expresses each element as normalized gaps (in terms of distance away from thresholds in Z), the matrix satisfies monotonicity[10]. Matrix G1 can be used directly to compute an equally weighted deprivation index for each zip code – This would simply be the sum of each individual row in matrix G1.

Optional: Exploratory Factor Analysis – Determining Index weightages

The construction of our deprivation index presently uses 5 sub-indicators as described in Table 1 above. Typically, Multidimensional Poverty Indices often contain more than 10 indices of interest. In that event, exploratory factor analysis is often employed to group sub-indicators together, or to discover the index weightages based on empirical commonalities between individual components. One strength is that it can identify components with strong correlations, but the weakness is that it implicitly assumes that only those components with strong correlations are relevant.

The general outline of factor analysis are as follows:

1. Extract factors
    a. Compute factor contributions using principal component analysis
2. Determine appropriate number of factors
    a. Kaiser criterion[11] – Pick only factors associated with eigenvalues of 1.0 or higher.
    b. Scree plot – Identifies inflexion point at which the adding of one more factor yields very little information on the overall index.
3. Rotate solution to obtain simple structural model in which each dimension(s) loads mostly to one factor. This step allows for ease of interpretation.

In our study, while we developed the code to compute factor loading, we did not use the associated weighted deprivation index in our visualization. This is because we only have 5 dimensions, and the scree plot of factor weights indicates only one factor associated with an eigenvalue of above 1.0. Nonetheless, we expect this segment to be useful if this project is expanded to include a much larger number of dimensions in the future.

Data Visualization

After the cleaning of the data and construction of the deprivation index, we made a series of

---

[10] A monotonic function mathematically preserves or reverses a given order. This attribute is desirable because it ensures that the ordering of individuals or regions based on their level of deprivation is consistent and unambiguous.

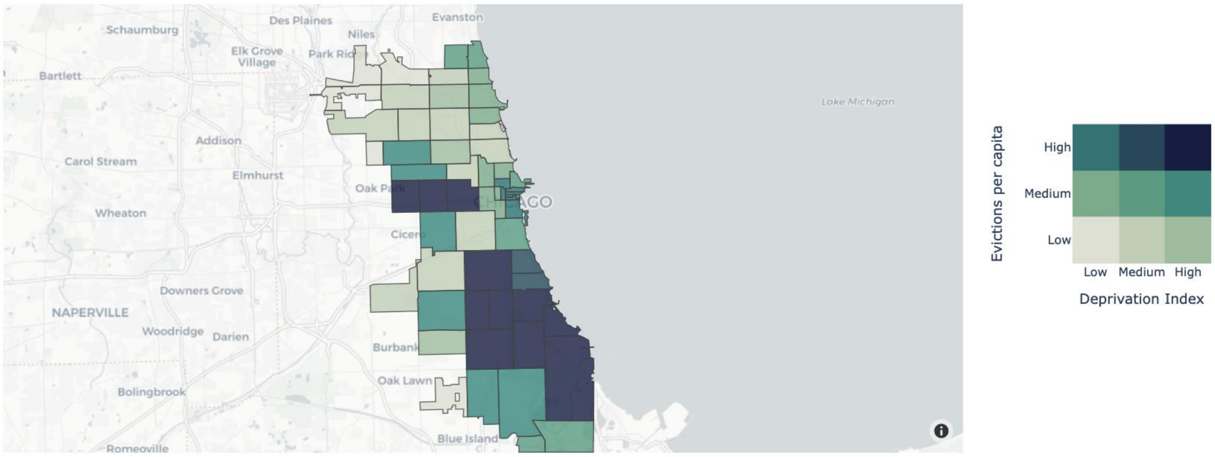
[11] The rationale is that each factor should not explain less than the equivalent of any of the individual dimensions that were included in the construction of the matrix.

visuals to describe the patterns in the data in Chicago and summarize our findings. Below are screenshots of the interactive dashboard we create:

We first show individual maps of evictions per capita and our deprivation index for the Chicago zip codes.

We also put both of these variables into a singular bivariate map, to show their joint distribution in Chicago zip codes.

How does neighborhood deprivation relate to evictions?



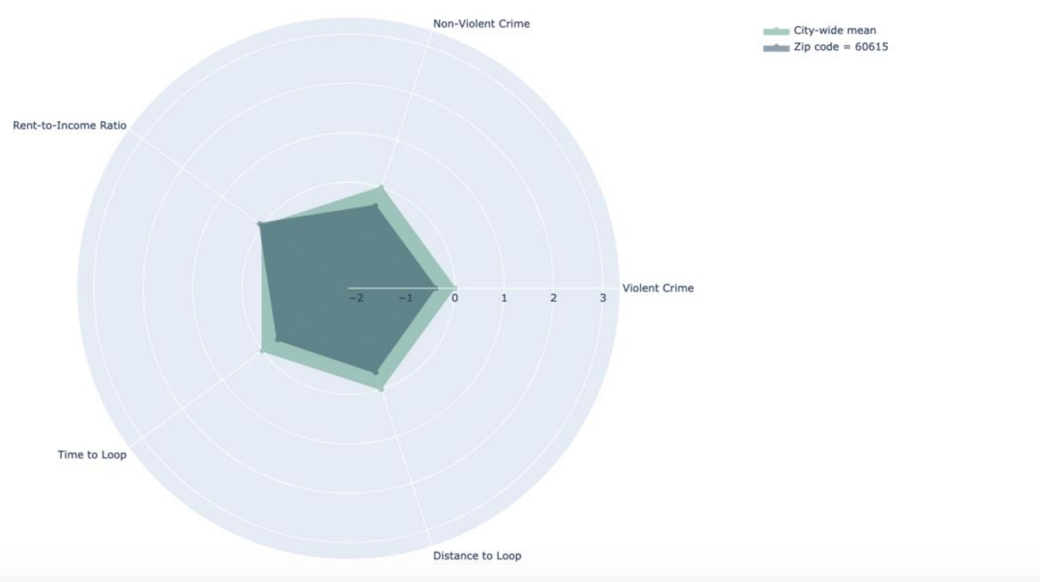
We also compare the deprivation index indicators for a single zip code against the city average. The user can select any zip code from the drop-down menu.

Comparison of Attributes by Zip Code to City Averages

The graph below compares the indicators which compose the deprivation index of an individual zip code to the overall average of all zip codes. All indicators are normalized such that their average is 0 and their standard deviation is 1. Doing this normalization makes the units for each of the indicators equivalent and allows them to be compared against each other. Higher values for each indicator indicate higher levels of deprivation; lower values for each indicator indicate lowers levels of deprivation.

Use the drop-down menu to select a zip code to compare to the city average:

60615



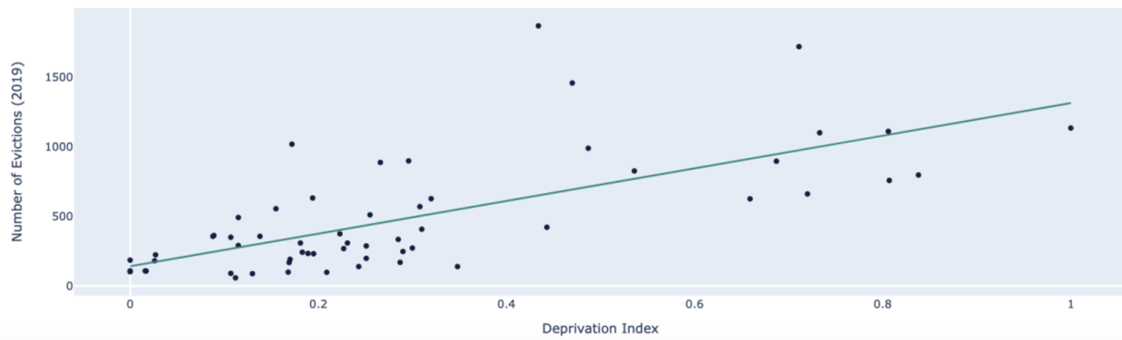
Finally, we compare the distribution of the deprivation index and its indicators to the number of evictions. In these graphs, each dot is a Chicago zip code.

Comparison of Evictions to Key Deprivation Indicators

The scatter plot below compares the distribution of the deprivation index to the distribution of evictions. Each dot in the scatter plot represents the values for a zip code. Use the drop-down to further compare the] the distribution of the individual indicators which compose the index against evictions All indicators are normalized such that their average is 0 and their standard deviation is 1. We overlay an OLS prediction line to visualize the linear relationship between the two variables.

Select the deprivation index or indicator to compare against per-capita evictions:

Deprivation Index ×



Conclusions

We find that our constructed deprivation index and the number of evictions are highly correlated at the zip code level. However, we are not able to determine the causal direction of this relationship from our analysis. Furthermore, some of the indicators which compose the deprivation index (time to the Loop and distance to the Loop) are not correlated with number of evictions. Future iterations of the deprivation index might exclude these indicators.